

I-027

興味領域情報を利用した映像インデクシングのためのカット画像自動分類

Classification of Cut Pictures for Video Indexing using Region of Interest

木村 雅之[†]
Masayuki KIMURA

山内 真樹[†]
Masaki YAMAUCHI

1 はじめに

HDD/DVD レコーダをはじめとするデジタル家電の急速な普及に伴い、個人であっても大量の映像コンテンツを所有、あるいはアクセスすることが一般的になりつつある。このような「コンテンツ爆発」とも呼べる状態においては、ユーザーに対して視聴・検索・あるいは編集における利便性をいかに提供するかが課題となる。たとえばテレビ番組に対しては、メタデータ(番組の内容を説明する情報)を利用することで、意味のあるまとまりごとに自動でチャプターを付与し、ボタン一つで見たいシーンを頭出しできるようにするなどといった視聴支援技術が今後重要になると考えられる。

これらの技術を実現するためには、基礎となるメタデータをいかにして付与するか、特に画像、音響特徴の解析によって自動的にメタデータを付与するための技術が必要不可欠である。しかし、一般的にメタデータ付与には高度なメディア理解技術が必須であり、これが大きな技術的障壁となっている。

一方で、例えばバラエティ番組において、スタジオのカットとビデオ映像のカットを、画像特徴レベルで区別することができれば、スタジオからビデオ映像への切り替わりの位置が分かり、結果として各シーンからの頭出し機能が実現できる。このように、番組中の各カットの類似度を判定してカテゴリ分けを行い、何らかのルールを元にまとまりのあるシーンを抽出する方法は、高度なメディア理解を介さずに処理を行うことができる点で有効であると考えられる [1]。

本論文ではカット画像に対して興味領域、すなわち人間の注目を集める何らかのものが存在する、と判断される領域を抽出し、興味領域における画像特徴を元にカット画像を分類する手法を提案する。これによって、画像内の被写体の位置ずれなどの影響を排した、より人間の感覚に近い分類が可能となる。

2 興味領域情報を利用したカット画像自動分類

2.1 興味領域の抽出

興味領域を抽出することは、認知心理学における視覚注意(Visual Attention)の位置および範囲の決定問題の一種と捉え

ることができる。例えば Itti ら [2] は色や輝度などを元に画像内の誘目度の大小を表す saliency map を作成し、それを元に視覚注意とその時間的変遷をモデル化している。Itti らのモデルはあくまで視覚注意の時間的な振る舞いを意識したものであるが、彼らが利用した saliency map を元に興味領域や被写体そのものの抽出が試みられている [3]。

これに対して、筆者ら [4] は画像のエッジ強度と近傍画素間の色相差をベースに誘目度を算出し、それを元に興味領域を抽出するアルゴリズムを独自に考案した。同アルゴリズムの適用例を図 1 に示す。



図 1 興味領域抽出の例

2.2 カット画像分類アルゴリズム

類似画像の分類・検索はすでに多くのアルゴリズムが提案されている。しかし、その多くは画像全体、あるいは一定のブロックごとの特徴量の比較を行っており、「人間がどこを見て類似性を判断しているか」を考慮した例は多くはない。本手法は、2.1 章で述べた興味領域を元に、以下の 2 つの特徴量を利用する。

1. 興味領域の画像中での位置
2. 興味領域の内外一定範囲の画像特徴

特に 2. について、具体的には興味領域の境界線から内外一定範囲の画素について、エッジ強度 (16 段階) と色相のヒストグラム (有彩色 12 段階 + 無彩色 3 段階) をそれぞれ取得する。なお、ヒストグラムは参照した画素数で正規化しておく。

カット画像の分類は、最短距離法に基づくクラスタリングによって行う。この際、2 つの特徴量を個別に用いることで、2 段階のクラスタリングを行う。まずはじめに、興味領域間の距離 d を距離尺度としてクラスタリングを行う。 d は式 (1) の通りである (図 2 参照)。

$$d = d_l^2 + d_r^2 + d_b^2 + d_g^2 \quad (1)$$

興味領域間の距離を比較することで、カット画像の大まかな構図を判定する効果が見込まれる。

次に、1 回目のクラスタリングでメンバ数が多いクラスタを順に一定個数選び、そのクラスタに属する画像に対して、ヒス

[†] 松下電器産業株式会社 AV コア技術開発センター, AV Core Technology Development Center, Matsushita Electronic Industrial Co., Ltd.

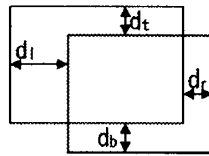


図2 興味領域間の距離

トグラムを元に距離 d' (式(2)参照)を設定し、2回目のクラスターリングを行う。

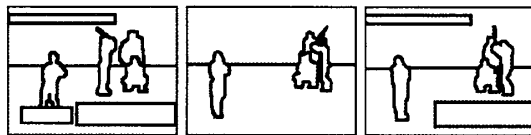
$$d' = \sum_{i=0}^{15} \{ \min(h_{edge,x}(i), h_{edge,y}(i)) \} + \sum_{j=0}^{14} \{ \min(h_{hue,x}(j), h_{hue,y}(j)) \} \quad (2)$$

但し、 h_{edge}, h_{hue} はそれぞれエッジ強度、色相のヒストグラムを表す。ヒストグラムを比較することで、カット画像の前景および背景の類似度を判定する効果が見込まれる。

両者を組み合わせることによって、先に述べた「人間がどこを見て類似性を判断しているか」を大まかに考慮しつつ、注目している部分の近傍の特徴を捉えることで、より人間の直感に近い分類が可能となる。

3 テレビ番組への適用結果

本手法を実際のテレビ番組に適用した例を以下に示す。なお、カット画像の検出は人手で行った。



(a) 野球



(b) バラエティ番組

図3 分類結果(概略図)

図3(a)は野球番組への本手法の適用結果の概略図である。特にピッチングシーンについて、テロップの有無や打者の立ち位置などの影響を受けず、安定して同カテゴリと判定している。また、図3(b)は、バラエティ番組への本手法の適用結果である。特に背景の一致度が高いスタジオ内の映像のカットが同カテゴリとして判定されている。

なお、上記野球番組における処理時間は、カット画像数1,413に対して、Pentium4 3GHzのPCで約45秒であった。ただし、この処理時間は、カット画像の読み込み、興味領域の算出、特徴量の計算、クラスターリングにかかる時間の総計である。

4 インデクシングの例

インデクシングの例として、本手法によって分類されたクラスターを元に、チャプタを作成する。チャプタ作成の方法として、例えば以下のような例が考えられる。

- 最大のクラスターに属するカットが現れるタイミングにチャプタを付与する
- 上位いくつかのクラスターに属するカットを時間軸上でまとめ、その切り替わりのタイミングにチャプタを付与する(図4参照)

前者は野球に代表されるような、特定の構造の画像が繰り返し現れ、それが映像の内容的な切れ目でもあるような番組に対して有効な方法である。一方、後者は例えばスタジオ内の映像とスタジオ外の映像が繰り返されるような、長い周期での映像の切り替わりを捉える方法である。図4の×印が上位のクラスターに属するカットを表し、その疎密が映像の切り替わりに対応していると考えられる。

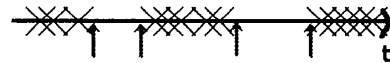


図4 チャプタ付与の例

5 まとめ

本論文では、画像から人の注目を集める興味領域を抽出し、その位置および領域の境界近傍の画像特徴を元に、放送コンテンツのカット画像の分類を行う手法について述べた。同時に、本手法を実際の放送コンテンツに適用し、分類結果に対してチャプタを付与する際の方法について検討を行った。今後の課題としては、クラスターリングアルゴリズムの高速化ならびに省資源化、チャプタ付与のルール作成、および多数のコンテンツへの本手法の適用による評価、などが挙げられる。

参考文献

- [1] 青木, “映像対話検出によるテレビ番組コーナ構成高速解析システム”, 電子情報通信学会論文誌 D-II Vol.88 No.1 pp.17-27, 2005-1
- [2] L. Itti, C. Koch, “A saliency-based search mechanism for overt and covert shifts of visual attention”, Vision Research 40, pp.1489-1506, 2000-6
- [3] Y. Hu, L. Chia, D. Rajan, “Region-of-Interest based Image Resolution Adaptation for MPEG-21 Digital Item”, ACM MultiMedia 2004, 2004-10
- [4] 木村, 山内, “画像からの興味領域抽出の一手法”, 画像の認識・理解シンポジウム MIRU2004 論文集 I, pp.93-94, 2004-8